On the Accuracy of Dense Fisheye Stereo

Johannes Schneider Cyrill Stachniss Wolfgang Förstner

***Abstract*— Fisheye cameras offer a large field of view, which is important for several robotics applications as a larger field of view allows for covering a large area with a single image. In contrast to classical cameras, however, fisheye cameras cannot be approximated well using the pinhole camera model and this renders the computation of depth information from fisheye stereo image pairs more complicated. In this work, we analyze the combination of an epipolar rectification model for fisheye stereo cameras with existing dense methods. This has the advantage that existing dense stereo systems can be applied as a black-box even with cameras that have field of view of more than 180 deg to obtain dense disparity information. We thoroughly investigate the accuracy potential of such fisheye stereo systems using image data from our UAV. The empirical analysis is based on image pairs of a calibrated fisheye stereo camera system and two state-of-the-art algorithms for dense stereo applied to adequately rectified image pairs from fisheye stereo cameras. The canonical stochastic model for sensor points assumes homogeneous uncertainty and we generalize this model based on an empirical analysis using a test scene consisting of mutually orthogonal planes. We show (1) that the combination of adequately rectified fisheye image pairs and dense methods provides dense 3D point clouds at 6-7 Hz on our autonomous multi-copter UAV, (2) that the uncertainty of points depends on their angular distance from the optical axis, (3) how to estimate the variance component as a function of that distance, and (4) how the improved stochastic model improves the accuracy of the scene points.**

I. Introduction

The ability to observe a large area in front of a camera is important for several applications. As a result of that, monocular and stereo cameras with a large field of view are becoming more and more popular. Examples include surveillance systems, unmanned aerial vehicles, see Figure 1 and [22], [28] or humanoid robots [4], [17], [19]. Camera systems with a large field of view mainly use wide-angle or fisheye lenses, mirrors, multiple cameras or rotating cameras. Fisheye lenses are an attractive choice as they offer several advantages in the image acquisition process such as a large field of view, robust mechanics and availability of very small form factors. They record a large field of view at each time of exposure, they avoid difficult to calibrate mirrors, they are comparably robust from a mechanical point of view and are available at small form factors. Using pairs of fisheye cameras allows to capture a large field of view stereoscopically, which is useful for monitoring the space around the sensors e.g. for obstacle avoidance. In contrast to classical cameras, however, fisheye lenses do not follow a perspective projection and cannot be approximated well using the pinhole camera model. This holds especially for cameras with a field of view of more than 180° and this often prevents the use of methods that assume a perspective projection model. This paper targets at computing dense stereo information from fisheye cameras and provides a detailed analysis of the quality of the recoverd 3D points with respect to the fisheye specific light projection on the image planes.

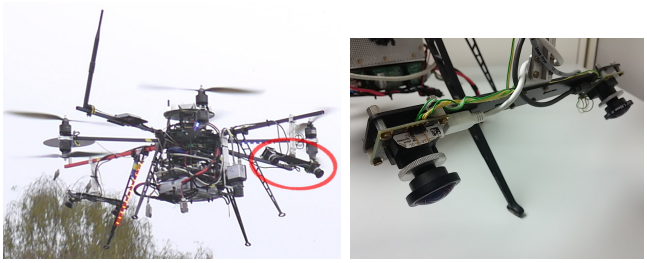

Fig. 1. Our UAV (left) equipped with fisheye stereo cameras with an opening angle of 185°. The paper describes how dense fisheye stereo can be computed based on existing methods for perspective cameras and analyzes the accuracy of the obtained point cloud from a theoretical and experimental perspective. The overall system runs at 6-7 Hz on our copter and provides 3D point clouds including information about its accuracy to improve reconstruction.

Traditional approaches to stereo vision rely on sparse points for which the 3D position is estimated through triangulation. The availability of sparse depth data only leads to more difficult object segmentation [31], scene understanding, or obstacle detection tasks. Thus, there is an increasing interest in semi-dense and dense reconstruction approaches [5] with applications in transportation systems [31], autonomous cars [**?**], or unmanned aerial vehicles [26].

A central task in sparse as well as dense stereo methods is to identify correspondences between the image pairs. By exploiting the epipolar geometry, we can reduce the 2D search problem to a simpler 1D problem. Depending on the used projection model for calibration and rectification, this 1D space corresponds to a straight line in a perspective projection or to a more complicated curve, e.g. a circular line in a stereographic projection [10]. Most systems for dense stereo assume that this 1D space is a straight line in the image, sometimes even that this line corresponds to a row in the image. This assumption can prevent the direct use of wide-angle or fisheye cameras with out-of-the-box dense stereo algorithms.

The contribution of this paper is an approach for re-using existing dense stereo methods with fisheye cameras. For this, we follow the approach of Abraham and Förstner [1] and generate virtual stereo image pairs that can then be used with existing dense stereo methods that assume the epipolar

The authors are with the Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn, Germany. This work has partially been supported by the DFG-Project FOR 1505 Mapping on Demand and by the EC under contract number H2020-ICT-644227-FLOURISH. We would gratefully thank Uwe Franke for providing SGM.

lines to correspond to a row in the image. This has the great advantage that highly optimized existing dense stereo methods can be applied as a black-box without modifications even with cameras that have field of view of more than 180°. In this paper, we consider semi-global matching (SGM) by Hirschmüller [14] and efficient large-scale stereo (ELAS) by Geiger et al. [9] but our approach is not restricted to these methods. Using the obtained disparity image, we derive a dense 3D point cloud together with the uncertainty of each single 3D point. We provide a detailed accuracy analysis of the obtained dense stereo results. This requires a realistic stochastic model for the disparities of the matched image points. Core of the paper therefore is a rigorous variance component estimation that optimally estimates the variance of the disparity at a point as a function of the distance of that image point to the image center and thus allows to predict the accuracy of the 3D points. We evaluate the significance of the improved stochastic model on scene reconstruction.

II. Related Work

Stereo matching is a large research area and a substantial number of algorithms for identifying stereo correspondence has been proposed. An good overview is given by Scharstein and Szeliski [25]. Over the last decade, more dense stereo and reconstruction methods have been developed. Popular approaches include semi-global matching by Hirschmüller [14] and efficient large-scale stereo by Geiger et al. [9]. Most of the dense stereo techniques have been designed for perspective cameras and cannot directly deal with the input of fisheye cameras. The idea of combining fisheye camera calibration and epipolar rectification for stereo computations goes back to Abraham and Förstner [1], which presented a method that can be seen as a specialization of the work by Pollefeys [24]. Esparza et al. [6] use a modified version of the epipolar rectification model to allow for wide stereo bases and largely disaligned optical axes. They apply epipolar rectification only on the overlapping image parts which allows the fast matching of detected keypoints along the image rows. Other rectification approaches exists, for example for binocular cylindrical panoramic images [15], which limits the vertical field of view and do not lead to epipolar images.

A review of fisheye projection models is given by Abraham and Förstner [1]. The work also provides an approach to calibrate fisheye stereo camera systems. Tommaselli et al. [30] showed that all the projection models in [1] are equally suitable to model fisheye cameras by comparing the residuals in 3D reconstruction after calibration. Fu et al. [8] determine the intrinsic and extrinsic parameters of a camera system that can consist of many overlapping fisheye cameras by using a wand with three collinear feature points and provide a toolbox online. Calibration approaches for a camera system with non-overlapping fisheye cameras are described in [27] and [11], both approaches use bundle adjustment without the need of fiducial markers.

Wang et al. [32] gives a formula to calculate the loss of spatial resolution of a fisheye camera with increasing distance to the image center. Their approach improves the image quality in regions with small spatial resolution using compressive sensing assuming a equi-distant projection model [34], but they do not provide a rigorous statistical analysis of their results.

Computing stereo information from fisheye cameras has also been investigated by other researchers. For example, Kita [16] analyzes dense 3D measurements obtained with a fisheye stereo camera pair with perfect calibration observing the workspace of a humanoid robot. Herrera et al. [12] propose a strategy for obtaining a disparity map from hemispherical stereo images captured with fisheye lenses in forest environments. To support the dense stereo process, they segment and classify the textures in the scene and consider only those matches belonging to the same class. Also Moreau et al. [20] address dense 3D point cloud computation using fisheye stereo pairs using epipolar curves with a unit sphere model. Arfaoui et al. [2] use cubic spline functions to model tangential and radial distortions in panoramic stereovision systems to simplify stereo matching. They also provide the mathematical relationship between matches to determine 3D point locations. Compared to our approach, neither Kita, Herrera et al., Moreau et al. nor Arfaoui et al. can exploit existing dense stereo implementations as a black box. Furthermore, they do not provide a detailed analysis of the accuracy of their results.

In addition to the dense stereo approaches, several new dense 3D reconstruction systems have been proposed in recent years, for example, Dense Tracking and Mapping by Newcombe et al. [21] or the approach by Stühmer et al. [29] that computes a dense reconstruction using variational methods. The simultaneous optimization of dense geometry and camera parameters is possible but a rather computationally intensive task [3]. In order to deal the computational complexity for real-time operation, semi-dense approaches are becoming increasingly popular, e.g. [5] for even monocular cameras.

Visual 3D reconstruction received also quite some attention in the context of light-weight UAV systems over the past few years. Especially in this application, light-weight sensors with a large field of view are attractive due to the strong payload limitations. For example Pizzoli et al. [23] propose a dense reconstruction approach for UAVs. They build upon a single perspective camera and their approach combines Bayesian estimation and convex optimization performing the reconstruction on a GPU at framerate. Related to that, combinations of perspective monocular cameras on an indoor UAV and RGB-D cameras on a ground vehicle have been used for simultaneous localization and mapping tasks aligning the camera information with dense ground models [7]. In contrast, our methods allows for using dense stereo methods with fisheye camera used on UAVs and provides an estimate of the accuracy of the returned point-cloud as illustrated in the motivating example in Figure 1.

III. Dense Stereo Methods for Perspective Cameras

In our work, we consider two popular dense stereo methods for computing a dense depth reconstruction given a stereo pair.

These two methods are efficient large-scale stereo (ELAS) [9] and semi-global matching (SGM) [14]. Both have been designed for calibrated perspective cameras and the output of both methods is a disparity image.

ELAS [9] and its implementation LibELAS compute disparity maps from rectified stereo image pairs and are robust against moderate illumination changes. ELAS provides a generative probabilistic model for stereo matching, which allows for dense matching using small aggregation windows. The Bayesian approach builds a prior over the disparity space by forming a triangulation on a set of robustly matched correspondences, so-called support points. ELAS applies a maximum a-posteriori estimation scheme to compute the disparities given all observations in the other image which are located on the given epipolar line. This yields an efficient algorithm with near real-time performance that also allows for parallelization.

Semi-Global matching [14] aims at combining local and global techniques in order to obtain an accurate, pixel-wise matching at comparably low computational requirements. It uses mutual information as the matching cost for corresponding points and the global radiometric difference is modeled in a joint histogram of corresponding intensities. An extension of SGM relies on the Census matching cost. Census is slightly inferior to mutual information, if there are only global radiometric differences, but it has been shown to outperform the mutual information in the presence of local radiometric changes and thus is beneficial in most real-world applications [13].

SGM uses a global cost function that penalizes small disparity steps, which are often part of slanted surfaces, less than real discontinuities. The cost function is optimized similarly to scan-line optimization and it finds an efficient solution for the 1D case. The key idea in SGM is to perform this computation along eight straight line paths ending in the pixel considering symmetry from all directions. Each path encodes cost for reaching the pixel with a certain disparity. For each pixel and each disparity, the costs are summed over the eight paths and at each pixel, the disparity with the lowest cost is chosen.

IV. Dense Fisheye Stereo and its Accuracy

This section describes our approach to obtain a dense 3D point cloud together with its uncertainty information using a stereo camera with fisheye lenses. The following two subsections introduce the equi-distance model. It describes the fisheye-specific light projection and the epipolar rectification model for fisheye cameras proposed in [1] that makes common dense stereo methods applicable. The third subsection describes how we compute the dense 3D point cloud with its uncertainty through variance propagation using the disparity information.

A. Fisheye Model

The fisheye specific projection from a 3D ray to a 2D image point can be described using the so-called equi-distance model, which is a reasonable first-order approximation for the

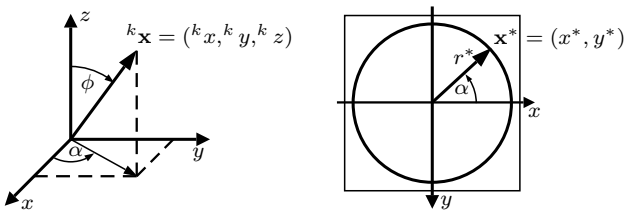

Fig. 2. Left: Camera ray ${}^k\boldsymbol{x}$ specified by angles ϕ and α in camera frame with optical axis z. Right: Relation between direction angles and conditioned image point coordinates $\boldsymbol{x}^*$.

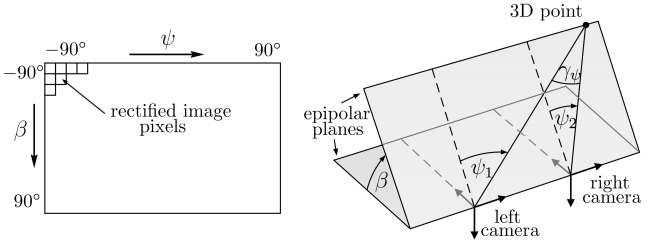

Fig. 3. The projection of the epipolar planes inside the rows according to (3). Each pixel coordinate of rectified image corresponds directly to the angles β and ψ.

intrinsically non-perspective projection of fisheye lenses [33]. The equi-distance projection model projects a 3D camera ray ${}^k\boldsymbol{x} = [{}^k x, {}^k y, {}^k z]^\mathsf{T}$ in the camera reference frame (indicated by superscript k), whose orientation is specified by the two angles ϕ and α as depicted in Figure 2, into a 2D position

$$\boldsymbol{x}^* = \begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} \frac{\operatorname{atan2}({}^k r, {}^k z)}{{}^k r}\, {}^k x \\ \frac{\operatorname{atan2}({}^k r, {}^k z)}{{}^k r}\, {}^k y \end{bmatrix} = \sin\phi \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} \quad (1)$$

with ${}^k r = \sqrt{{}^k x^2 + {}^k y^2}$. Note that the projection is radial symmetric in relation to the optical axis. The radial distance in the conditioned image $r^* = \sqrt{x^{*2} + y^{*2}} = \phi$ only depends on the angle ϕ between the 3D ray ${}^k\boldsymbol{x}$ and the optical axis and becomes a monotonously increasing function, which allows for a field of view even larger than 180°.

The relation of conditioned image point $\boldsymbol{x}^*$ to its unconditioned coordinates is given by $\boldsymbol{x}' = c\ \boldsymbol{x}^* - \boldsymbol{h}$ with the principal point $\boldsymbol{h} = [h_x, h_y]^\mathsf{T}$ and the principal distance c obtained by camera calibration e.g. according to [1]. Given a 2D point $\boldsymbol{x}^*$, the inverse transformation of (1) into a 3D camera ray reads as

$${}^k\boldsymbol{x} = \left[{}^k x, {}^k y, {}^k z\right]^\mathsf{T} = \left[\frac{\sin r^*}{r^*} x^*,\ \frac{\sin r^*}{r^*} y^*,\ \cos r^*\right]^\mathsf{T}. \quad (2)$$

In Section IV-C, we will use this model to propagate the positional uncertainty of an observed image point to its corresponding camera ray. Note that we have not introduced additional parameters for lens distortion and assume them to be negligible small after proper calibration.

B. Epipolar Rectification

In a camera pair with two projection centers, all epipolar planes intersect at the baseline. Despite ideal properties of the stereo cameras like parallel optical axis, the introduced equi-distant projection model does not lead to images where each 3D point is projected into the same row in both

cameras, thus the epipolar lines are curved. To obtain parallel epipolar lines such that the vertical disparity vanishes and the correspondence search can be reduced to a one-dimensional search along the image rows, we use the epipolar rectification model proposed by [1]. We exploit the concept of a virtual camera to achieve a rectification for the image pair that is independent of the real projection system and leads to ideal properties: identical interior orientation with no distortions, no camera rotation and a baseline in one axis direction. The epipolar equi-distance rectification model projects the epipolar planes to the same image row in both images.

The projection function is given by

$$\boldsymbol{x}^* = \begin{bmatrix} \operatorname{atan2}\left({}^k x, \sqrt{{}^k y^2 + {}^k z^2}\right) \\ \operatorname{atan2}\left({}^k y, {}^k z\right) \end{bmatrix} = \begin{bmatrix} \beta \\ \psi \end{bmatrix} \quad (3)$$

where the coordinates of the conditioned image point $\boldsymbol{x}^*$ correspond directly to the angles ψ and β that describe the ray to the observed 3D point as shown in Figure 3: β characterizes the pitch angle of each epipolar plane and ψ characterizes the projection inside the epipolar plane, i.e. the image row.

For image rectification principal distance c and principal point $\boldsymbol{h}$ from calibration can be used. Given an image pixel position $\boldsymbol{x}'$ in the rectified image the corresponding angles are than obtained by the relation $[\beta, \psi]^\mathsf{T} = (\boldsymbol{x}' - \boldsymbol{h})/c$.

The transformation from conditioned image position $\boldsymbol{x}^*$ into a ray direction ${}^k\boldsymbol{x}$ with unit length is given by

$${}^k\boldsymbol{x} \;=\; [\sin x^*,\; \cos x^* \sin y^*,\; \cos x^* \cos y^*]^\mathsf{T} \;. \quad (4)$$

C. 3D Point Cloud with Uncertainty

We derive the 3D point coordinates with their uncertainty through variance propagation given an image point with its disparity information. Let $\Sigma_{\boldsymbol{x}'\boldsymbol{x}'}$ describe the positional uncertainty of the image point $\boldsymbol{x}' = [x', y']^\mathsf{T}$ in the *unrectified image* given by

$$\Sigma_{\boldsymbol{x}'\boldsymbol{x}'} = \operatorname{Diag}([\sigma_{x'}^2,\, \sigma_{y'}^2]). \quad (5)$$

For the fisheye lenses, we use the equi-distant camera model according to Section IV-A. Using the principal distance c and principal point $\boldsymbol{h}$ from calibration, we obtain the conditioned image coordinates $\boldsymbol{x}^*$ with their covariance matrix $\Sigma_{\boldsymbol{x}^*\boldsymbol{x}^*}$ as

$$\boldsymbol{x}^* = (\boldsymbol{x}' - \boldsymbol{h})\,/c \quad \text{and} \quad \Sigma_{\boldsymbol{x}^*\boldsymbol{x}^*} = \operatorname{Diag}([\sigma_{x'}^2,\, \sigma_{y'}^2])/c^2. \quad (6)$$

This yields the corresponding camera ray ${}^k\boldsymbol{x}$ according to (2) and its covariance matrix through variance propagation

$$\Sigma_{{}^k\boldsymbol{x}\,{}^k\boldsymbol{x}} = \boldsymbol{J}_1 \Sigma_{\boldsymbol{x}^*\boldsymbol{x}^*} \boldsymbol{J}_1^\mathsf{T} \quad (7)$$

with

$$\boldsymbol{J}_1 = \begin{bmatrix} \frac{\sin(r^*)y^{*2} + \cos(r^*)x^{*2}r^*}{(x^{*2}+y^{*2})^{3/2}} & \frac{(\cos(r^*)r^* - \sin(r^*))y^*x^*}{(x^{*2}+y^{*2})^{3/2}} \\ \frac{(\cos(r^*)r^* - \sin(r^*))y^*x^*}{(x^{*2}+y^{*2})^{3/2}} & \frac{\cos(r^*)y^{*2}r^* + \sin(r^*)x^{*2}}{(x^{*2}+y^{*2})^{3/2}} \\ -\frac{\sin(r^*)x^*}{r^*} & -\frac{\sin(r^*)y^*}{r^*} \end{bmatrix}. \quad (8)$$

Given the previously defined rectification, we obtain the angles ψ and β from a ray ${}^k\boldsymbol{x}$ according to (3) and for the covariance matrix follows

$$\Sigma_{\left[\begin{smallmatrix}\beta\\ \psi\end{smallmatrix}\right]} = \boldsymbol{J}_2 \Sigma_{{}^k\boldsymbol{x}\,{}^k\boldsymbol{x}} \boldsymbol{J}_2^\mathsf{T} \quad (9)$$

with

$$\boldsymbol{J}_2^\mathsf{T} = \begin{bmatrix} \frac{\sqrt{{}^k y^2 + {}^k z^2}}{{}^k x^2 + {}^k y^2 + {}^k z^2} & 0 \\ \frac{-{}^k x\,{}^k y}{\sqrt{{}^k y^2 + {}^k z^2}({}^k x^2 + {}^k y^2 + {}^k z^2)} & \frac{{}^k z}{{}^k y^2 + {}^k z^2} \\ \frac{-{}^k x\,{}^k z}{\sqrt{{}^k y^2 + {}^k z^2}({}^k x^2 + {}^k y^2 + {}^k z^2)} & \frac{-{}^k y}{{}^k y^2 + {}^k z^2} \end{bmatrix}. \quad (10)$$

As the corresponding camera rays do intersect in one point (as β is identical for both rays), we can determine its coordinates easily. Let s be the distance from the left camera along the camera ray ${}^k\boldsymbol{x}$ to the unknown 3D point $\boldsymbol{p} = s\,{}^k\boldsymbol{x}$. Camera ray ${}^k\boldsymbol{x}$ can be derived with β and ψ according to (4). To compute s, we use the angles β and ψ and the ψ-disparity γ_ψ given with the image coordinates of corresponding points, see also Figure 3. Note that the apical angle, i.e. the intersection angle, complies with the disparity angle

$$\gamma_\psi = \gamma_{x'}/c \quad (11)$$

with the measured disparity $\gamma_{x'}$ in the epipolar rectified image and the principal distance c used for this rectification. This can be shown using the angular sum $\gamma_\psi = 180° - \psi_1' - \psi_2'$ with the interior angles $\psi_1' = 90° - \psi$ and $\psi_2' = 90° + \psi - \gamma_\psi$.

Exploiting the law of sines, we obtain

$$s = b\,\frac{\sin(90° + \psi - \gamma_\psi)}{\sin\gamma_\psi} = b\,\frac{\cos(\psi - \gamma_\psi)}{\sin\gamma_\psi}, \quad (12)$$

with b being the base line, which leads to the 3D coordinates of the point $\boldsymbol{p}$ as

$$\boldsymbol{p}(\psi, \beta, \gamma_\psi) = b\,\frac{\cos(\psi - \gamma_\psi)}{\sin\gamma_\psi} \begin{bmatrix} \sin\psi \\ \cos\psi\sin\beta \\ \cos\psi\cos\beta \end{bmatrix}. \quad (13)$$

With the vector $\boldsymbol{q} = [\psi, \beta, \gamma_\psi]^\mathsf{T}$, the covariance matrix of $\boldsymbol{p}$ is obtained through

$$\Sigma_{\boldsymbol{p}\boldsymbol{p}} = \boldsymbol{J}_3 \operatorname{Diag}([\Sigma_{\left[\begin{smallmatrix}\beta\\ \psi\end{smallmatrix}\right]},\, \sigma_{\gamma_\psi}^2])\, \boldsymbol{J}_3^\mathsf{T} \quad \text{with} \quad \boldsymbol{J}_3 = \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{q}}. \quad (14)$$

The individual elements of $\boldsymbol{J}_3$ in (14) are the partial derivatives of (13) and are best obtained using a symbolic calculation toolbox such as Mathematica or Maple and are not shown here due to the sake of brevity.

V. Improved stochastic Observation Model

We start with a *standard stochastic model* for the observed entities. The sensor coordinates of the images points are assumed to be identically and independently distributed $\mathbb{D}([x_i'; y_i']) = \sigma_x^2\, \boldsymbol{I}_2$ and the disparities are assumed to have the same variance $\mathbb{D}(\gamma_\psi) = \sigma_{\gamma_\psi}^2$. Due to the properties of the optics, we can expect in a first approximation that the accuracy of the sensor coordinates depends on the angle ϕ between the viewing direction and the direction to the scene point.

In order to determine this dependency, we observe planar surfaces in a scene and analyze the residuals using a robust version of variance component analysis leading to a refined or *improved stochastic model* for the observation's variances. Using a stochastic model which is closer to reality should

Fig. 4. Left images: Stereo camera with fisheye lenses and highly textured and mutually orthogonal planes $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ used for variance analysis. Upper right images: Stereo image pair. Lower right: Image pair after epipolar rectification. Note that all epipolar lines of the left and right image are in the same row.

lead to better estimates of the plane's parameters. We will check this empirically by analyzing orthogonal planes.

A. Variance Analysis

Classical estimation procedures assume the covariance matrix $\Sigma_{\boldsymbol{ll}}$ of the $n = 1, ..., N$ observations to be known up to an unknown variance factor, where $\boldsymbol{l}$ refers to the observations. Thus, the stochastic model is assumed to be $\Sigma_{\boldsymbol{ll}} = \sigma_0^2 \Sigma_{\boldsymbol{ll}}^a$, where $\Sigma_{\boldsymbol{ll}}^a$ is an approximation for the covariance matrix, and the unknown variance factor σ_0^2 is assumed to be one. Based on a Gauss-Markov model of the form

$$p(\boldsymbol{l}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{a}, \sigma_0^2 \Sigma_{\boldsymbol{ll}}^a) \tag{15}$$

with the Jacobian $\boldsymbol{A}$ and U unknown parameters, we obtain the ML-estimate

$$\widehat{\boldsymbol{x}} = \Sigma_{\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}} \boldsymbol{A}^\mathsf{T} \Sigma_{\boldsymbol{ll}}^{-1} (\boldsymbol{l} - \boldsymbol{a}) \tag{16}$$

with the covariance matrix

$$\Sigma_{\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}} = (\boldsymbol{A}^\mathsf{T} \Sigma_{\boldsymbol{ll}}^{-1} \boldsymbol{A})^{-1}. \tag{17}$$

With the estimated residuals $\widehat{\boldsymbol{v}} = \boldsymbol{A}\widehat{\boldsymbol{x}} + \boldsymbol{a} - \boldsymbol{l}$ and the redundancy $R = N - U$, we have the unbiased estimated variance factor

$$\widehat{\sigma}_0^2 = \widehat{\boldsymbol{v}}^\mathsf{T} \Sigma_{\boldsymbol{ll}}^{-1} \widehat{\boldsymbol{v}} / R \qquad \text{with} \qquad \sigma_{\widehat{\sigma}_0} = \sqrt{2/R}\, \sigma_0 \,. \tag{18}$$

For an *improved stochastic model*, we now assume that the variances of the observations follow the model

$$\Sigma_{\boldsymbol{ll}} = \sum_{j=1}^{J} \sigma_j^2 \Sigma_j^a \tag{19}$$

with known approximate covariance matrices and unknown variance factors, also called variance components, σ_j^2. In our case, we assume

$$\sigma_{l'_n}^2 = \sigma_1^2 + \sigma_2^2 \phi_n^{2p} \tag{20}$$

i.e. the noise of the sensor coordinates is a sum of a constant noise term $\underline{n}_1$ with $p(n_1) = \mathcal{N}(0, \sigma_{01}^2)$ and a noise term $\underline{n}_2$ proportional to the p-th power ϕ_n^p of the angle ϕ_n referring to the n-th observation, thus $p(n_2) = \mathcal{N}(0, \sigma_{02}^2)$. As we will illustrate in the experimental evaluation through the analysis of the variance factors computed for different angles ϕ, this models describes the noise in relation to ϕ well.

This leads to the two covariance matrices

$$\Sigma_1^a = \boldsymbol{I}_N \qquad \text{and} \qquad \Sigma_2^a = \text{Diag}([\phi_n^{2p}]) \,. \tag{21}$$

With the weight or precision matrix $\boldsymbol{W}_{\boldsymbol{ll}} = \Sigma_{\boldsymbol{ll}}^{-1}$ of the observations and the covariance matrix $\Sigma_{\widehat{\boldsymbol{v}}\widehat{\boldsymbol{v}}} = \Sigma_{\boldsymbol{ll}} - \boldsymbol{A}^\mathsf{T} \Sigma_{\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}} \boldsymbol{A}$, the general and the specific expressions for the estimated variance components are

$$\widehat{\sigma}_j^2 = \frac{\widehat{\boldsymbol{v}}^\mathsf{T} \boldsymbol{W}_{\boldsymbol{ll}} \Sigma_j^a \boldsymbol{W}_{\boldsymbol{ll}} \widehat{\boldsymbol{v}}}{\text{tr}\,(\boldsymbol{W}_{\boldsymbol{ll}} \Sigma_j^a \boldsymbol{W}_{\boldsymbol{ll}} \Sigma_{\widehat{\boldsymbol{v}}\widehat{\boldsymbol{v}}})} \,. \tag{22}$$

In our case, this simplifies to the relations

$$\widehat{\sigma}_1^2 = \frac{\sum_n w_n^2 \widehat{v}_n^2}{\sum_n w_n^2 \sigma_{\widehat{v}_n}^2} \quad \text{and} \quad \widehat{\sigma}_2^2 = \frac{\sum_n w_n^2 \widehat{v}_n^2 \phi_n^{2p}}{\sum_n w_n^2 \sigma_{\widehat{v}_n}^2 \phi_n^{2p}} \,. \tag{23}$$

The estimated variance factors lead to an updated covariance matrix of the observations as in (19) and we apply the estimation procedure iteratively until convergence.

B. Orthogonality Improvement

The improved stochastic model should lead to better estimates of the plane's parameters. In case of mutually orthogonal planes the angle ω between the estimated normal directions should get closer to 90° than when using the classic stochastic model.

Estimating the orthogonal planes N times using different stereo images leads to $n = 1, ..., N$ deviations $\omega_n - 90°$. The empirical variance $\widehat{\sigma}_\omega^2 = \frac{1}{N} \sum_n (\omega_n - 90°)^2$ and the theoretical variance σ_ω^2 derived from covariance matrix $\Sigma_{\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}}$ of both estimated planes should (a) indicate an higher precision than when using the classical model and (b) confirm empirically the plausibility of the stochastic model if $\widehat{\sigma}_\omega$

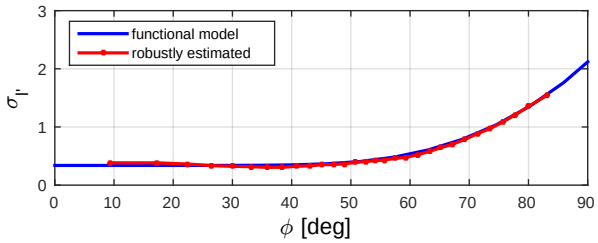

Fig. 5. The red dots show 30 robust estimates for standard deviation $\widehat{\sigma}_{l'}$ using the residuals of narrow ranges of ϕ, the blue line shows the estimated functional model of $\widehat{\sigma}_{l'}$ over angle ϕ.

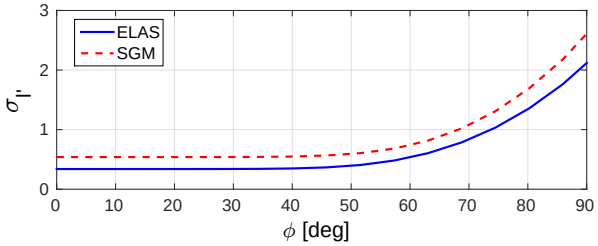

Fig. 6. Estimated standard deviation $\widehat{\sigma}_{l'}$ over angle ϕ using disparities from ELAS (solid, blue) and SGM (dashed, red).

and σ_ω comply with the relative accuracy of (18), i.e. if $\widehat{\sigma}_\omega/\sigma_\omega \approx \sqrt{2/N}$ holds.

VI. Experimental Evaluation

The goal of this experimental evaluation is to illustrate that dense fisheye stereo can be achieved and to investigate the accuracy of dense stereo with fisheye cameras using the epipolar rectification model. For the evaluation, we use a stereo camera with a basis of 20 cm and Lensagon BF2M14420 fisheye lenses with a field of view of 185°. We calibrated the stereo camera by estimating the interior and relative orientation according to [1] using the epipolar equi-distant rectification model. For epipolar rectification, we use a camera constant of $c = 200$ pixel to keep most of the image content in 752×480 images. After rectification the disparity between corresponding points is limited to the same image row, see Figure 4.

A. Variance Analysis

For the first two sets of experiments, we use three highly textured and mutually orthogonal planar surfaces, see Figure 4, for evaluating the variance analysis described in Section V. To analyze the accuracy of the observations in dependency of the angle ϕ, we capture the three planar surfaces under 30 different poses such that the planes are visible over a broad spectrum of ϕ. For each image pair, we use ELAS and SGM to determine dense disparity information. For ELAS, we use the default settings for robotic environments as well as the default settings for SGM.

For each pixel with disparity information, we obtain the coordinates of a 3D point $\boldsymbol{p}$ in camera frame using (13). We compute the covariance matrix $\Sigma_{\boldsymbol{pp}}$ according to (14) using a standard stochastic model with identically and independently distributed image points and disparities $\sigma_{\gamma_{x'}} = \sigma_{x'} = \sigma_{y'} = 1$ pixel. We then estimate for each of the 30 captured stereo pairs the three normal directions of the three planes $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ in a robust RANSAC procedure using the covariance weighted residuals of the points to identify outliers.

We directly obtain the residual for every inlier point by computing its distance to the plane in the direction of the normal directions whereas each point belongs either to $\mathcal{A}_1$, $\mathcal{A}_2$ or $\mathcal{A}_3$. Using all transformed points from all 30 stereo pairs with their angle ϕ from the optical axis of the camera, we estimate the best plane and update the variance factors $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$ according to (20). This is done iteratively, updating the estimated variance factors and scaling the covariance matrices $\Sigma_{\boldsymbol{pp}}$ according to the point specific angle ϕ.

We use the exponent $p = 8$ in (20) as this model describes the robust determined variances of the residuals over ϕ best. The dots on the red line in Figure 5 indicate the obtained standard deviations using a robust version of variance factor estimation. For this, we determine the variance of the residuals for narrow ranges of ϕ, by partitioning the set of all ϕ_n in 30 equally sized bins. For each bin, we use $\widehat{\sigma}_{l'} = 1.48\,\mathrm{MAD}$ with the median absolute difference (MAD) of the residuals to obtain a robust estimate for the standard deviation, see [18] for details. The blue line in Figure 5 shows the estimated functional model of $\widehat{\sigma}_{l'}$ in (20) in dependency of ϕ with $p = 8$, which is close to the 30 determined standard deviations.

Figure 6 shows the estimated standard deviation $\widehat{\sigma}_{l'}$ after convergence in dependency of ϕ using the disparities obtained with ELAS (as in Figure 5) and SGM. Both curves have the similar shape and the difference amounts about 0.2 pixel. As this figure shows, measurements having an angle ϕ less than 40° from the optical axis have the highest and nearly constant precision of 0.3 and 0.5 pixel. Beyond 40° the precision degrades revealing the substantially smaller precision of the disparities towards the image borders. By knowing this function, we can now exploit this information in the improved stochastic model.

B. Orthogonality of Planes

Table I shows the empirically derived mean and standard deviation of all derived 30 angles. The improved stochastic model that considers the influence of ϕ on the precision of the 3D points leads in all cases to smaller deviations from orthogonality and is therefore closer to reality. The empirically derived standard deviations $\widehat{\sigma}_w$ of the angles between the planes confirm a higher precision. The theoretic standard deviation σ_ω that can be obtained given our model is on average 0.424 (ELAS) and 0.676 (SGM) times smaller using the estimated variance factors we obtained in practice. The quotient $\widehat{\sigma}_\omega/\sigma_\omega$ is throughout in the range of $\sqrt{2/30}$ around one hence the proposed improved stochastic model of the observation process complies with the empirical results.

C. Application Examples

Finally, we want to illustrate that the described approach is able to build dense 3D point clouds in real world situations. Therefore, we show results from an indoor and outdoor scene.

Figure 7 shows the point cloud derived from a stereo image taken in an office with the fisheye stereo camera described before. The disparity information is obtained with ELAS on the epipolar rectified images. The color of each point

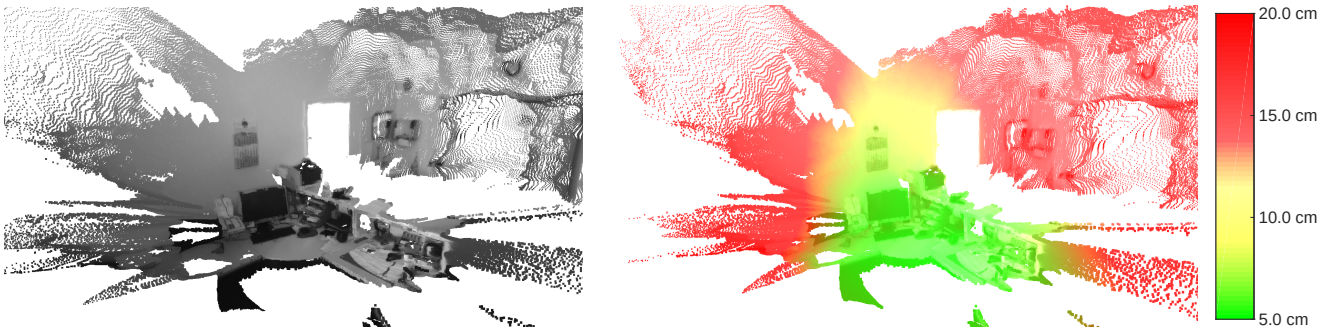

Fig. 7. Left: Point cloud obtained with disparity information from ELAS. The intensity values correspond to the image content of the left image; Right: Color according to the position accuracy of 3D point, which ranges from 5 cm to 20 cm (green high, red low accuracy).

	$\angle(\mathcal{A}_1, \mathcal{A}_2)$	$\angle(\mathcal{A}_1, \mathcal{A}_3)$	$\angle(\mathcal{A}_2, \mathcal{A}_3)$
ELAS			
classical	$89.69° \pm 1.49°$	$90.23° \pm 0.89°$	$89.74° \pm 0.96°$
improved	$89.93° \pm 0.63°$	$90.02° \pm 0.65°$	$90.04° \pm 0.36°$
SGM			
classical	$89.95° \pm 1.29°$	$89.92° \pm 1.19°$	$89.80° \pm 0.58°$
improved	$89.93° \pm 0.76°$	$89.94° \pm 0.84°$	$89.87° \pm 0.30°$

TABLE I: Empirically derived mean and std. deviation $\widehat{\sigma}_\omega$ of the 30 estimated angles ω between two orthogonal planes using the disparity information from ELAS/SGM and the classical or the improved stochastic model.

corresponds in the left image to the recorded pixel intensity. In the right image, the intensities are overlaid with the theoretical precision obtained with the estimated variance model. The color spectrum goes from green for points with highest precisions of about 5 cm over yellow to red for points with lowest precision up to 20 cm. Highest precision is achieved for points on the desk as the angles of intersecting rays from both cameras (γ_ψ) are high and the angle ϕ is small in the center of the image (see Figure 7). Points on the wall behind the desk have smaller disparity angles γ_ψ thus less precision (yellow). The precision decreases with angle ϕ and leads to more noisy 3D points more distant to the camera axis (red).

In the last example, we compare the point cloud of a agricultural surface obtained with a fisheye stereo image taken from our copter with a reference point cloud. To compare both point clouds a rigid body transformation was estimated using corresponding 3D points. The reference point cloud has a point accuracy of about 1 cm and was obtained by bundle adjustment and a subsequent densification using high resolution images taken with high-end equipment. The stereo camera is tilted with $45°$ towards the ground and the dense depth information from the fisheye stereo image pair is shown in Figure 8. It depicts the colored reference point cloud overlayed with the fisheye cloud. The different color encoding shows the absolute error for each point and the histogram illustrates the error distribution. As can be seen, the quality of dense stereo information decays away from the optical axis as the stochastic model predicts. Areas with high errors also have a high theoretical uncertainty.

D. Remarks

The processing of a fisheye stereo image pair which includes the rectification, disparity determination and mapping of the 3D points takes in our ROS implementation 150 ms per image pair using the default robotics parameters in ELAS thus enables real-time applications. Thus, we can process stereo

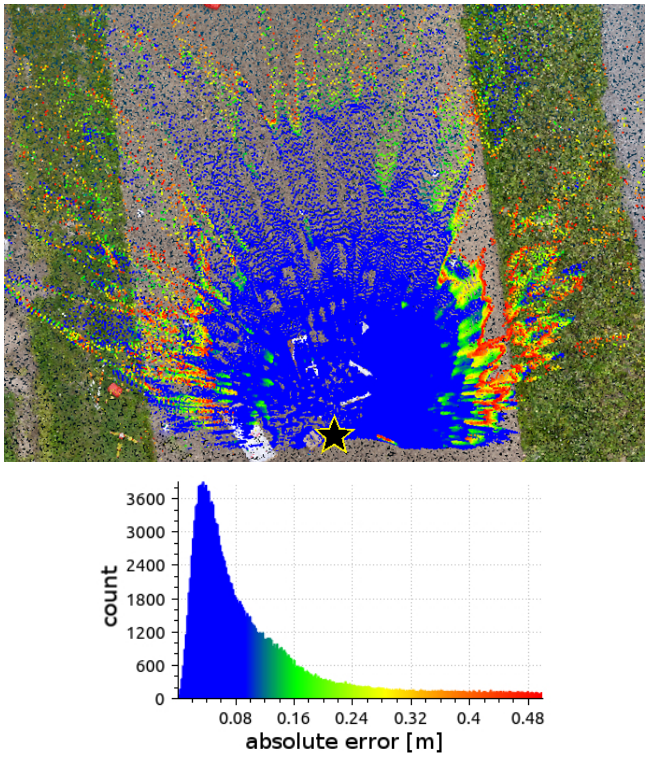

Fig. 8. Point cloud obtained from a single 752×480 pixel fisheye image pair overlayed with a reference point cloud. The histogram illustrates the distribution of the absolute distances between nearest neighbours encoded with different colors. The black star marks the position of the copter at the time of exposure.

images with 6-7 Hz, which is suited for online operation in several application scenarios.

Our experiments suggest that in combination with fisheye epipolar rectification, ELAS and SGM can both be used for dense fisheye stereo and both methods perform similarly. The precision of the 3D points decreases with angle ϕ. For $\phi > 40°$, the precision drops substantially and leads to more noisy 3D points more distant to the optical axis. This information can be exploited within the observation model. In our experiments, the improved model for the noise in the observations yields a better estimate than the standard model. The theoretic standard deviation σ_ω is on average between 0.42 and 0.68 times smaller than the ones obtained experimentally.

VII. Conclusions

In this paper, we analyzed an approach to exploit existing dense stereo methods with wide-angle and fisheye cameras that have a field of view of more than 180°. By conducting fisheye calibration and epipolar rectification beforehand, we can use existing state-of-the-art dense stereo methods as a black box. We thoroughly investigated the accuracy potential of such a fisheye stereo approach and derived an estimate of the uncertainty of the obtained 3D point cloud. We furthermore generalized the canonical stochastic model for sensor points based on an empirical analysis. We showed (1) that adequately rectified fisheye image pairs and dense methods provides dense 3D point clouds at 6-7 Hz, (2) that the uncertainty of image points depends on their angular distance from the center of symmetry, (3) how to estimate the parameters of a variance component model, and (4) how the improved stochastic model for the observations influences the accuracy of the 3D points. Please note that our method is not limited to a specific fisheye stereo camera system. The limitations of the disparity determination correspond directly to the limitations of the used dense stereo algorithm, e.g. in structureless environments.

Acknowledgment

The authors would like to thank Christian Eling and Lasse Klingbeil from the University of Bonn for their support during the UAV experiments.

References

[1] S. Abraham and W. Förstner. Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 59(5):278–288, 2005.

[2] A. Arfaoui and S. Thibault. Mathematical model for hybrid and panoramic stereovision systems: panoramic to rectilinear conversion model. *Applied Optics*, 54(21):6534–6542, 2015.

[3] M. Aubry, K. Kolev, B. Goldluecke, and D. Cremers. Decoupling photometry and geometry in dense variational camera calibration. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2011.

[4] M. Bennewitz, C. Stachniss, W. Burgard, and S. Behnke. Metric localization with scale-invariant visual features using a single perspective camera. In *European Robotics Symposium*, pages 143–157, 2006.

[5] J. Engel, J. Sturm, and D. Cremers. Semi-dense visual odometry for a monocular camera. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, pages 1449–1456, 2013.

[6] J. Esparza, H. Helmle, and B. Jähne. Wide base stereo with fisheye optics: A robust approach for 3d reconstruction in driving assistance. In *Proc. of the German Conf. on Pattern Recognition (GCPR)*, volume 8753 of *Lecture Notes in Computer Science*, pages 342–353, 2014.

[7] C. Forster, M. Pizzoli, and D. Scaramuzza. Air-ground localization and map augmentation using monocular dense reconstruction. In *Proc. of the Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 3971–3978, 2013.

[8] Q. Fu, Q. Quan, and K.-Y. Cai. Calibration of multiple fish-eye cameras using a wand. *IET Computer Vision*, 9(3):378–389, 2014.

[9] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Proc. of the Asian Conf. on Computer Vision (ACCV)*, volume 6492 of *Lecture Notes in Computer Science*, pages 25–38, 2010.

[10] J. Heller and T. Pajdla. Stereographic rectification of omnidirectional stereo pairs. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1414–1421, 2009.

[11] L. Heng, M. Bürki, G.H. Lee, P. Furgale, R. Siegwart, and M. Pollefeys. Infrastructure-based calibration of a multi-camera rig. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2014.

[12] P.J. Herrera, G. Pajares, M. Guijarro, J.J. Ruz, and J.M. Cruz. A stereovision matching strategy for images captured with fish-eye lenses in forest environments. *IEEE Sensors Journal*, 11:1756–1783, 2011.

[13] H. Hirschmüller. Semi-global matching – motivation, developments and applications. In *Proc. of the Photogrammetric Week*, pages 173–184, 2011.

[14] H. Hirschmller. Stereo processing by semi-global matching and mutual information. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*, 30(2):328–341, 2008.

[15] H. Ishiguro, M. Yamamato, and S. Tsuji. Omni-directional stereo. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*, 14(2):257–262, 1992.

[16] N. Kita. Dense 3d measurement of the near surroundings by fisheye stereo. In *Proc. of the Conf. on Machine Vision Applications*, pages 148–151, 2011.

[17] N. Kita. Direct floor height measurement for biped walking robot by fisheye stereo. In *IEEE Intl. Conf. on Humanoid Robots*, pages 187–192, 2011.

[18] K.R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer Berlin, 2 edition, 1999.

[19] D. Maier, C. Stachniss, and M. Bennewitz. Vision-based humanoid navigation using self-supervised obstacle detection. *The Int. Journal of Humanoid Robotics (IJHR)*, 10, 2013.

[20] J. Moreau, S. Ambellouis, and Y. Ruichek. Equisolid fisheye stereovision calibration and point cloud computation. In *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XL-7/W2, pages 167–172, 2013.

[21] R.A. Newcombe, S. Lovegrove, and A.J. Davison. Dtam: Dense tracking and mapping in real-time. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2011.

[22] M. Nieuwenhuisen, D. Dröschel, J. Schneider, D. Holz, T. Läbe, and S. Behnke. Multimodal obstacle detection and collision avoidance for micro aerial vehicles. In *Proc. of the European Conf. on Mobile Robotics (ECMR)*, pages 7–12, 2013.

[23] M. Pizzoli, C. Forster, and D. Scaramuzza. Remode: Probabilistic, monocular dense reconstruction in real time. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pages 2609–2616, 2014.

[24] M. Pollefeys, R. Koch, and L. Van Gool. A simple and efficient rectification method for general motion. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, volume 1, 1999.

[25] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Intl. Journal of Computer Vision (IJCV)*, 47:7–42, 2002.

[26] K. Schmid, P. Lutz, T. Tomic, E. Mair, and H. Hirschmüller. Autonomous vision-based micro air vehicle for indoor and outdoor navigation. *Journal of Field Robotics (JFR)*, 31:537–570, 2014.

[27] J. Schneider and W. Förstner. Bundle adjustment and system calibration with points at infinity for omnidirectional camera systems. *Photogrammetrie – Fernerkundung – Geoinformation (PFG)*, 4:309–321, 2013.

[28] J. Schneider and W. Förstner. Real-time accurate geo-localization of a mav with omnidirectional visual odometry and gps. In *ECCV Workshop: Computer Vision in Vehicle Technology (CVVT)*, volume 8925 of *Lecture Notes in Computer Science*, pages 271–282, 2014.

[29] J. Stühmer, S. Gumhold, and D. Cremers. Real-time dense geometry from a handheld camera. In *Proc. of the Annual Symposion of the German Association for Pattern Recognition (DAGM)*, pages 11–20, 2010.

[30] A.M.G. Tommaselli, J. Marcato Jr, M.V.A. Moraes, S.L.A. Silva, and A.O. Artero. Calibration of panoramic cameras with coded targets and a 3d calibration field. In *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XL-3/W1, pages 137–142, 2014.

[31] W. van der Mark and D.M. Gavrila. Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):38–50, 2006.

[32] W. Wang, H. Xiao, W. Li, and M. Zhang. Enhancement of fish-eye imaging quality based on compressive sensing. *Optik – Intl. Journal for Light and Electron Optics*, 126(19):2050–2054, 2015.

[33] Y. Xiong and K. Turkowski. Creating image-based vr using a self-calibrating fisheye lens. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 237–243, 1997.

[34] C. Xu and X. Peng. Fish-eye lens rectification based on equidistant model. In *Proc. of the Intl. Conf. on Information Technology and Applications (ITA)*, pages 163–166, 2014.